Light-Scattering Studies of Protein Solutions: Role of Hydration in Weak Protein-Protein Interactions

A. Paliwal, D. Asthagiri, D. Abras, A. M. Lenhoff, and M. E. Paulaitis
Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland

ABSTRACT We model the hydration contribution to short-range electrostatic/dispersion protein interactions embodied in the osmotic second virial coefficient, B_2, by adopting a quasi-chemical description in which water molecules associated with the protein are identified through explicit molecular dynamics simulations. These water molecules reduce the surface complementarity of highly favorable short-range interactions, and therefore can play an important role in mediating protein-protein interactions. Here we examine this quasi-chemical view of hydration by predicting the interaction part of B_2 and comparing our results with those derived from light-scattering measurements of B_2 for staphylococcal nuclease, lysozyme, and chymotrypsinogen at 25°C as a function of solution pH and ionic strength. We find that short-range protein interactions are influenced by water molecules strongly associated with a relatively small fraction of the protein surface. However, the effect of these strongly associated water molecules on the surface complementarity of short-range protein interactions is significant, and must be taken into account for an accurate description of B_2. We also observe remarkably similar hydration behavior for these proteins despite substantial differences in their three-dimensional structures and spatial charge distributions, suggesting a general characterization of protein hydration.

INTRODUCTION

Early efforts to describe the osmotic second virial coefficient, B_2, for protein solutions involved applications of standard colloidal models with the proteins represented by idealized geometries, such as spheres or ellipsoids (1–4). These models were then fit to experimental data by balancing contributions from electrostatic and dispersion interactions using the relative magnitude of the two contributions as an adjustable parameter. More sophisticated treatments (1,5,6) included permanent and induced dipole moment interactions and charge fluctuation interactions. Despite reasonable quantitative agreement with observed trends in protein solution behavior, the physically unrealistic parameters that were derived by data fitting, such as those for the Hamaker constant, remained worrisome (1–3,6,7). Reasons cited for the unrealistic parameter values included the inherent approximations made in applying simple geometries to represent protein shape, and the inability to account for more complex effects, such as those arising from solvation (8). Nevertheless, characterizations of protein-protein interactions based on simple interaction potentials and idealized geometries have provided the basis for generalized phase diagrams of broad classes of proteins (9,10).

In light of these earlier efforts, it is remarkable that the importance of molecular shape and charge complementarity in governing protein solution thermodynamics was appreciated as early as the 1940, at a time when no protein crystal structures were available (11). Molecular descriptions of protein-protein interactions have, however, received little attention until only recently. Neal et al. (12) used more realistic molecular thermodynamic models of protein solutions in accounting for both protein shape and charge heterogeneity to reveal the decisive role played by highly complementary protein-protein orientations. Specifically, they showed that attractive electrostatics coupled with geometric complementarity could explain the ionic strength-dependence observed in B_2 measurements for chymotrypsinogen solutions as a function of pH. A similar approach was developed by Elcock and McCammon (13), who accounted for protein structure in a more elaborate model for electrostatic interactions to describe the pH-dependence of B_2.

A molecular thermodynamic model of protein solutions in which the configurational complementarity of protein-protein interactions plays a central role naturally leads to a consideration of the molecular nature of protein hydration. The view of hydration obtained from small-angle scattering experiments (14), high resolution crystal structures (15), and molecular simulations (16,17) is of a primary hydration layer around the protein that is ~3 Å thick and has a density 10–20% higher than bulk water density. A primitive treatment of protein hydration considers the excluded volume effect of this primary hydration layer on protein-protein interactions, which is captured in calculating B_2 by effectively increasing the excluded volume of the protein (18). However, simply accounting for protein shape in estimating the excluded

Submitted April 26, 2005, and accepted for publication June 8, 2005.

A. Paliwal and D. Asthagiri contributed equally to this work.

Address reprint requests to M. E. Paulaitis, E-mail: paulaitis.1@osu.edu; or E-mail: michaelp@jhu.edu.

M. E. Paulaitis's present address is Dept. of Chemical and Biomolecular Engineering, Ohio State University, Columbus, OH 43210.

D. Asthagiri's address is Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87544.

A. M. Lenhoff's address is Dept. of Chemical Engineering, University of Delaware, Newark, DE 19716.

0006-3495/05/09/1564/10 $2.00

doi: 10.1529/biophysj.105.065284

volume contribution to B_2 for globular proteins gives rise to a contribution that is greater than that for the sphere of equivalent volume (19). The magnitude of this difference is roughly equivalent to adding a uniform hydration layer 3 Å thick, which suggests that incorporating a uniform ''hydration'' layer as an additional excluded volume contribution merely corrects for simplifying the protein geometry. Moreover, surface roughness at the atomic level, in addition to protein shape, can impact the excluded volume contribution to B_2, and therefore, a dense hydration layer of uniform thickness and adjustable scattering length density may also account for the side-chain packing efficiency of amino acids at the protein-water interface (14). A more detailed treatment of protein hydration considers the underlying local hydration structure (20,21). In this view, there is a spatial distribution of hydration sites at the protein-water interface characterized by the interactions of water molecules with the local environment. Averaging local densities at these specific hydration sites over the primary hydration layer gives a density higher than that of bulk water. More importantly, though, a spatially heterogeneous distribution of hydration sites can alter the configurational complementarity of protein-protein interactions in the ensemble of highly favorable configurations that dominate B_2 calculations.

In a recent publication (22), we presented an analysis of osmotic second virial coefficients obtained by light-scattering from protein solutions. Two principal contributions to B_2 were identified: 1), the (ideal) Donnan contribution, which accounts for electroneutrality in a multicomponent solution of (poly)electrolytes; and 2), a contribution due to protein-protein interactions in the limit of infinite dilution. Distinguishing these separate contributions allowed us to model the interaction part of B_2 by molecular computations. In comparing our model predictions with measurements of B_2 for lysozyme, we found that long-range electrostatic interactions dominate the interaction part of B_2 at low ionic strength; however, short-range electrostatic/dispersion interactions with specific hydration are essential for an accurate description of B_2 derived from experiment. Specific hydration is accounted for in the model by adopting a quasi-chemical description in which we consider an ensemble of explicit water molecules that are taken to be strongly associated with the protein. The effect of including specific hydration is to reduce short-range attractive dispersion interactions by attenuating a number of highly complementary protein-protein configurations.

Here, we analyze the role of specific hydration in more detail. The criterion used in our treatment to identify strongly associated water molecules from molecular dynamics simulations is examined, and a more general characterization of hydration behavior for small globular proteins is undertaken in the context of this quasi-chemical description of hydration. We also report new results for B_2 obtained by light-scattering from staphylococcal nuclease solutions, and compare our model predictions for staphylococcal nuclease (SN), lysozyme (LYS), and chymotrypsinogen (CGA) to our experimental data as well as data reported in a previous study (6). An analysis of the interaction part of B_2 reveals differences in the solution thermodynamic behavior for CGA relative to SN and LYS that highlights the importance of evaluating the individual contributions to B_2.

THEORY

In this section, we briefly review our analysis of light-scattering from protein solutions and describe the model for protein-protein interactions as background for developing a thermodynamic treatment of specific hydration. Details of this analysis and the theoretical basis for our model are given elsewhere (22).

The starting point of our analysis is Stockmayer's expression for the turbidity of a solution due solely to the composition fluctuations (23), which requires partial derivatives of chemical potentials with respect to the molar concentrations of the solution components. We consider a solution consisting of a protein component, added salt component, and solvent (water), where a component is defined to be electrically neutral, although each electroneutral component may consist of charged constituents (24). The solvent is denoted by *0*, the salt component by *1*, and the protein component by *2*. For simplicity, the added salt is assumed to be NaCl, and the protein, P, carries a positive charge z. Therefore, the protein component is PCl_z. The concentrations of free HO^- and H^+ ions in solution are taken to be negligible in comparison (compare with Ref. 24). The resulting expression for turbidity, τ, is

$$\frac{H\psi_2^2}{\tau}\rho_2 \approx 1 + \left(\frac{z^2}{2\rho_1} + \beta_{22}\right)\rho_2 = 1 + 2B_2\rho_2, \qquad (1)$$

where ρ_i is the molar concentration of component i, $\psi_2 = \partial n/\partial\rho_2$ is the derivative of the refractive index of the solution with respect to the protein concentration, H is an optical constant, and the coefficient B_2 is identified as the osmotic second virial coefficient. The two contributions comprising B_2 are the so-called Donnan term, $z^2/2\rho_1$, and

$$\beta_{22} = (\partial\beta\tilde{\mu}_2/\partial\rho_2)_{T,p,\rho_1}, \qquad (2)$$

the nonideal contribution due to protein-protein interactions and the target of our molecular modeling efforts. Here $\tilde{\mu}_2$ is the excess chemical potential of the protein component, and $\beta = 1/kT$ with kT the thermal energy. In this analysis, we have neglected partial derivatives of $\tilde{\mu}_i$ with respect to ρ_1, which represent nonideal contributions arising from preferential partitioning of the salt ions in the vicinity of protein molecules that become important at high protein charge and/or high ionic strength (25). Also, for the protein concentrations and ionic strengths considered here $\rho_1 \gg \rho_2$, and we have made the physically reasonable assumption that $\psi_2 \gg \psi_1$ (24).

We adopt the Debye-Hückel model of protein-protein interactions in the limit of large protein-protein separations, which is taken as the protein surfaces separated by more than a Debye length, or $r > (a + \kappa^{-1})$, where a is the nominal diameter of the protein, and κ is given by the usual expression,

$$\kappa^2 = 4\pi\frac{\beta}{\epsilon}\sum_i \rho_i z_i^2, \tag{3}$$

where ϵ is the solution dielectric constant. In the limit $\rho_2 \rightarrow 0$, κ^2 is proportional to ρ_1, the ionic strength of the solution due to the added salt. Recognizing that the Donnan contribution is recovered from the Debye-Hückel model, we obtain the expression for the long-range contribution to β_{22},

$$\beta_{22,l} + \frac{z^2}{2\rho_1} = \frac{z^2}{2e\rho_1}\left[\frac{2+\kappa a}{1+\kappa a} - \frac{\kappa}{4e}\frac{\beta}{\epsilon}\left(\frac{z^2}{1+\kappa a}\right)^2\right], \tag{4}$$

where e is the base of the natural logarithm. The additional subscript l emphasizes that this contribution corresponds to just the long-range part of β_{22}.

For protein-protein interactions at separations between a and $(a + \kappa^{-1})$, we express the shape and orientation dependence of the potential of mean force (PMF) between protein molecules in solution explicitly, and write

$$\begin{aligned}\beta_{22,s} &= -\frac{1}{8\pi^2}\int_\Omega\int_a^{a+1/\kappa}(e^{-\beta w(r,\Omega)} - 1)r^2 dr d\Omega \\ &= -\frac{1}{8\pi^2}\int_\Omega I(\Omega)d\Omega = -4\pi\langle I\rangle_\Omega,\end{aligned} \tag{5}$$

for the short-range contribution to β_{22}. Here Ω is a collective variable for all Euler and polar angles describing the relative orientation of the two protein molecules, and $w(r, \Omega)$ is the PMF in that orientation as a function of separation, r. The term $I(\Omega)$ represents the contribution to $\beta_{22,s}$ of the particular orientation, Ω.

Our final expression for β_{22} is

$$\begin{aligned}\beta_{22} = \frac{7\pi}{6}a^3 &+ \frac{z^2}{2e\rho_1}\left[\frac{2+\kappa a}{1+\kappa a} - e - \frac{\kappa}{4e}\left(\frac{\beta}{\epsilon}\right)\left(\frac{z}{1+\kappa a}\right)^2\right] \\ &- \frac{1}{8\pi^2}\int_\Omega\int_a^{a+1/\kappa}(e^{-\beta w(r,\Omega)} - 1)r^2 dr d\Omega,\end{aligned} \tag{6}$$

where the first term on the right-hand side is the protein-excluded volume plus an approximation to the small contribution due to the protein-solvent PMF. The electrostatic contributions embodied in $w(r, \Omega)$ are calculated using an earlier model (12), as described in Asthagiri et al. (22). Dispersion interactions are modeled using a hybrid Lennard-Jones/Lifshitz-Hamaker approach also described earlier (26). The configurational integral in Eq. 6 is estimated by Monte Carlo sampling of the configurational space (27). A total of 10^4 configurations were generated for each calculation.

HYDRATION THERMODYNAMICS

Although the hybrid Lennard-Jones/Lifshitz-Hamaker approach captures the essential features of surface complementarity in intermolecular interactions, the effect of strongly associated water molecules is lost. In the quasi-chemical description of hydration (28), it is natural to view these water molecules as part of the protein. The solution thermodynamics is then described in terms of quasi-components comprised of the protein and associated water molecules immersed in a statistical field due to the exterior medium. Here, we identify strongly associated water molecules through explicit molecular dynamics (MD) simulations and retain the statistical description for the remainder.

We implement this quasi-chemical view of hydration as follows. First, a grid is constructed to fill a proximal volume around the protein such that each grid point defines a cubic volume 1 Å on a side. This proximal volume is taken to be within 3.5 Å of the heavy atoms comprising the protein surface. Selected grid points are then identified to represent those water molecules that are strongly associated with the protein. The selection of these grid points depends on the interaction of a water molecule with its local environment at each grid point relative to bulk water, which is expressed in terms of a chemical equilibrium constant for water association with the protein at that grid point, or equivalently,

$$\eta \equiv \ln(\rho/\rho_b) = -\beta\Delta\tilde{\mu}_{H_2O}, \tag{7}$$

where $\Delta\tilde{\mu}_{H_2O}$ is the difference in the excess chemical potential of water at the grid point and in bulk water, and ρ/ρ_b is the corresponding ratio of water densities. Eq. 7 provides the thermodynamic framework for selecting grid points to represent strongly associated water molecules based on MD simulations that supply the required densities. Grid points with $\eta > 0$ have, on average, water densities higher than the bulk density, and therefore are identified as locations for water molecules that associate with the protein. Our criterion for strong association is $\eta > 2$, which corresponds to a free energy of association that is more favorable than dissolution in bulk water by at least 2 kT. For SN, only 212 out of a total of 11,690 proximal grid points were observed to have $\eta > 2.0$. The observed numbers of proximal grid points with $\eta > 2.0$ for LYS and CGA were 135 out of 7855 and 267 out of 12,672, respectively. In each case, we assume that all grid points satisfying the criterion $\eta > 2.0$ are equivalent in terms of their respective affinities for water molecules, and as such, water molecules are retained at all of them in our model. However, we account for overpopulating the grid points with water molecules by decreasing the TIP3P Lennard-Jones ϵ-parameter for water molecules placed at these grid points by a factor of 3, which corresponds to the observation that there are roughly three times as many grid points as there are water molecules in the primary hydration layer that satisfy the criterion $\eta > 2.0$.

SN coordinates used in the simulation were obtained from the NMR-minimized average structure (29) (PDB ID: 1JOO). The NMR structure was preferred over the crystal structure (PDB ID: 1STN) because the latter has a total of 13 residues missing from loop regions at the C-and N-termini. Differences between the two structures other than these loop regions are minimal considering the thermal motion of the protein in water. The root mean-square deviation between corresponding α-carbons in the NMR and crystal structures is 2.3 Å compared to a root mean-square deviation of 1.7 Å in α-carbon positions computed from MD simulation (30).

SN was solvated by inserting the NMR structure into a cubic box 62 Å on a side containing 8000 TIP3P water molecules (31), and removing water molecules within 4 Å of the heavy atoms of the protein. A total of 6822 water molecules remained in the final system. The protein atoms were held fixed throughout the simulation since the model for β_{22} is based on a rigid protein structure. The system was initially relaxed by executing 20,000 steps of a steepest-descent minimization cycle. All simulations were carried out at 298 K and 1 bar using NAMD (32) with the CHARMM27 force field (33). The system temperature was held constant by applying the Langevin dynamics method to all nonhydrogen atoms with a damping coefficient of 1 ps^{-1}. The system pressure was maintained using a Nosé-Hoover Langevin piston (34) with a period of 200 fs and a decay of 100 fs. Periodic boundary conditions were imposed and electrostatic interactions were determined using the particle-mesh Ewald method (35) with a real-space cutoff of 12 Å. The same cutoff was used for nonbonded nonelectrostatic interactions. Water geometry was constrained by the SHAKE algorithm (36). Equilibration was carried out for 200 ps, followed by a production run of 2 ns with a time step of 2 fs. Water configurations were saved every 0.1 ps for further analysis.

MATERIALS AND METHODS

Wild-type SN was expressed and purified to ≥98% as determined by SDS-PAGE following the procedure of Shortle and Meeker (37). SN was flash-frozen in liquid nitrogen and stored at −80°C. For each experiment, fresh SN was thawed and centrifuged for at least 30 min to remove any misfolded/denatured protein and aggregates. SN solutions for the static light-scattering (SLS) measurements were prepared by either dialyzing extensively against the desired solution or by successive dilutions using Amicon Ultra Centrifugal Filter Tubes with 5000-MW cutoff (Amicon Plastics, Houston, TX).

NaCl (S9888, Sigma-Aldrich, St. Louis, MO) was used to adjust the ionic strength of the protein solutions. Buffer salts—sodium acetate (3470-01, J.T. Baker, Paris, KY), bis-Tris (156663, Sigma-Aldrich), and Tris (T-1503, Sigma-Aldrich)—were used to stabilize the pH at values of 5.0, 6.5, and 8.0, respectively. The pH was restricted to a maximum value of 8.0 to avoid SN aggregation near its pI (~10). The solution pH was measured using a Mettler Toledo MP220 pH meter (Mettler-Toledo AB, Stockholm, Sweden) and was adjusted by addition of small quantities of 0.1–0.5 M HCl (9535-33, J.T. Baker) and NaOH (3722-01, J.T. Baker). All SLS samples were prepared with filtered deionized water obtained from a Barnstead NANOpure UV water filter system (Barnstead International, Dubuque, IA). The buffer solutions were filtered with Whatman 20-nm inorganic filters (Whatman PLC, Brentford, UK) and were used to prepare stock solutions of SN at ~10 mg/ml and various solution conditions. The protein samples were filtered with Amicon Ultrafree MC centrifugal filter devices (Millipore, Billerica, MA) with 100-nm pore size before taking measurements. All glassware was first treated with detergent, stored overnight in HELLMAMEX II alkaline cleaning solution, and then washed thoroughly with filtered and deionized water, shortly before an experiment.

SLS data were collected at an angle of 90° on a Malvern 4700C system (Malvern Instruments, Malvern, Worcestershire, UK) equipped with a Lexel 95 Ar-ion laser (Lexel, Palo Alto, CA) operating at a wavelength of 488 nm, and a Malvern MULTI8 computing correlator (7032 CN). Toluene (TX 0735-6, EMD Chemicals, Gibbstown, NJ) was used as an index matching fluid in the glass container that held the sample cell. This container was also periodically cleaned with isopropyl alcohol during the course of the study to avoid dust affecting the measurements. A NesLab RTE-210 water bath (NesLab Instruments, Portsmouth, NH) was used to control the temperature at 25 ± 0.1°C by circulating water through the metal casing that encloses the glass container assembly. Benzene (HPLC grade; 27079, Sigma-Aldrich) was used as a calibration solvent to obtain the excess Rayleigh ratio of protein solutions.

In the SLS experiment, the Rayleigh ratio, R_θ, rather than turbidity, is usually measured. The Rayleigh ratio is related to the turbidity by $R_\theta = 3/8\pi \times \tau(1 + \cos^2\theta)$. Also, it is customary to use units of (g/vol) for the protein concentration. In such cases, $\rho_2 = c_2/M_w$, with c_2 in g/vol and M_w the molecular weight of the protein. With these units, Eq. 1 becomes

$$\frac{Kc_2}{R_{90}} = \frac{1}{M_w} + \frac{1000}{M_w^2}\left(\frac{z^2}{2\rho_1} + \beta_{22}\right)c_2, \tag{8}$$

where $K = 4\pi^2 n^2 (\psi_2)^2/\lambda^4 N_A$ is an optical constant, λ is the wavelength of incident light, N_A is Avogadro's number, and R_{90} is the excess Rayleigh ratio of the protein solution at $\theta = 90°$. A plot of the left side of Eq. 8 as a function of the protein concentration, c_2, gives $1/M_w$ as the intercept and $2B_2$ as the slope, where units of B_2 are mol ml/g^2. R_{90} for each sample was calculated by calibration with benzene during an experiment as

$$R_{90} = \left(\frac{I - I_S}{I_B}\right)\left(\frac{n}{n_B}\right)^2 R_{B,90}, \tag{9}$$

where I, I_S, and I_B are the scattered intensities at 90° for the protein sample, pure solvent, and benzene, respectively. The value n_B is the refractive index of benzene and $R_{B,90}$ is the absolute Rayleigh ratio of benzene at 90°, taken to be 38.6×10^{-6} cm^{-1} (6,38). A value of $\psi_2 = 0.2$ ml/g was found to best fit the light-scattering data for SN. This value lies within the narrow range reported for other globular proteins, such as chymotrypsinogen, lysozyme, and bovine serum albumin at $\lambda = 488$ nm (6,39). $R = R_{90}$ is implied in the discussion below.

RESULTS AND DISCUSSION

Light-scattering measurements of B_2

Fig. 1 shows plots of Kc_2/R as a function of SN concentration for several ionic strengths at pH 8.0. The same plots are shown in Fig. 2 with Kc_2/R adjusted for the Donnan contribution, $(1000c_2/M_w^2)z^2/2\rho_1$. The data in Fig. 2 are fit to quadratic functions of c_2 and limiting slopes are obtained by taking the derivative at $c_2 = 0$. The values of β_{22} derived from these slopes are given in Table 1. The weight-averaged molecular mass obtained from the intercepts is 17.6 ± 0.4 kDa, which is slightly greater than the molecular mass calculated from the amino-acid sequence of SN, 16.8 kDa. This result indicates the presence of higher molecular weight aggregates of SN in solution. Assuming these aggregates are

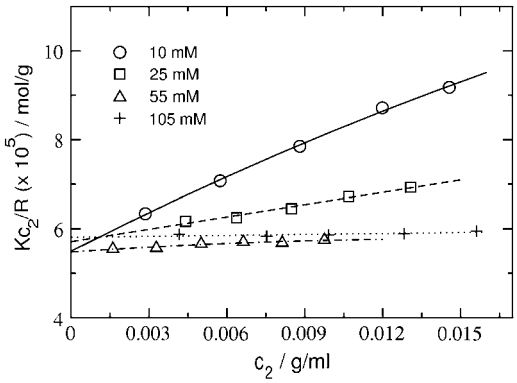

FIGURE 1 SLS plots for staphylococcal nuclease (compare with Eq. 8) at pH 8.0 and various ionic strength conditions. Lines represent quadratic fits to the data.

SN dimers, we estimate ~4.8% dimers by weight in solution, which is consistent with 1H NMR measurements (40) that find a small fraction of SN in dimers at the protein concentrations studied here. In comparison, our earlier light-scattering measurements for LYS (22) yielded within experimental uncertainties a single intercept corresponding to a molecular weight that was identical to the value calculated based on amino-acid sequence.

Table 1 summarizes the SLS results for SN at all solvent conditions. The values of $2B_2$ obtained directly from the limiting slope of Kc_2/R as a function of c_2 match those calculated as the sum of $z^2/2\rho_1$ and β_{22}. Therefore, only the latter values are reported in this table. We observe that B_2 is positive at all values of pH and decreases with ionic strength, consistent with the notion that repulsive electrostatic interactions are progressively screened with increasing ionic strength. We also find that β_{22} is negative at all values of pH, implying attractive nonideal protein-protein interactions, and becomes progressively less negative with ionic strength. Similar behavior was noted in our earlier light-scattering study of LYS solutions (22). The physical interpretation of this result, i.e.,

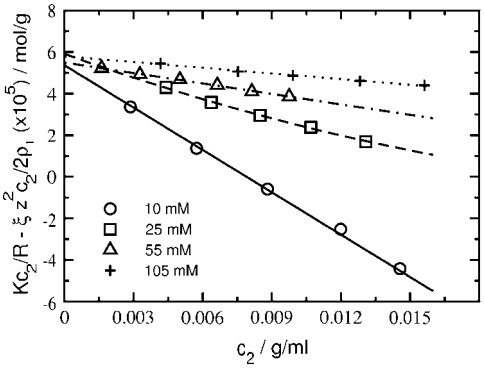

FIGURE 2 SLS plots for staphylococcal nuclease at pH 8.0 and various ionic strengths adjusted for the Donnan contribution, $\xi z^2 c_2/2\rho_1 (\xi = 1000/M_w^2)$. The curves are quadratic fits to the data; β_{22} is obtained from the slope in the limit $c_2 = 0$.

TABLE 1 SLS results for staphylococcal nuclease

pH	z	I	$2B_2$	$z^2/2\rho_1$	β_{22}
5.0	14	0.010	152.1	347.2	−195.1
5.0		0.030	38.1	115.7	−77.6
5.0		0.060	12.7	57.9	−45.2
5.0		0.110	8.5	31.6	−23.0
6.5	9	0.010	40.4	143.5	−103.1
6.5		0.025	13.7	57.4	−43.7
6.5		0.055	6.6	26.1	−19.5
6.5		0.105	2.1	13.5	−11.7
8.0	7.9	0.010	37.2	110.6	−73.4
8.0		0.025	10.4	44.2	−33.8
8.0		0.055	3.3	20.1	−16.8
8.0		0.105	1.1	10.5	−9.4

$B_2 \times 10^4$ and the contributing terms are given in units of mol ml/g^2 as is customary. I is the ionic strength (including buffer) in moles/liter. The value z is the charge carried by SN calculated using intrinsic pK_a values.

$$\beta_{22} = \frac{\partial \beta \tilde{\mu}_2}{\partial \rho_2} < 0, \tag{10}$$

is that adding protein in the limit $\rho_2 \rightarrow 0$ increases the ionic strength of the solution, which reduces the free energy of charging the distinguished protein molecule. The effect is to screen the long-range electrostatic interactions that dominate β_{22} at low ionic strength.

Modeling protein-protein interactions

The results for β_{22} and the contributing terms for SN are given in Table 2. Calculated values of β_{22} as a function of solution pH and ionic strength are also compared with those derived from our SLS experiments. We find that the agreement is reasonable and the correct ionic strength dependence is recovered. Moreover, the solution thermodynamic behavior is qualitatively similar to that reported earlier in our light-scattering study of protein-protein interactions in LYS solutions (22). Specifically, β_{22} is large and negative at low ionic strength and becomes progressively less negative with increasing ionic strength. In addition, the contribution to β_{22} from $\beta_{22,l}$ dominates, particularly at low ionic strength, and

TABLE 2 Modeling results for staphylococcal nuclease

pH	z	I	$\beta_{22,s} + 7\pi a^3/6$	$\beta_{22,l}$	$\beta_{22,calc}$	$\beta_{22,exp}$
5.0	14	0.010	35.5	−263.6	−228.1	−195.1
5.0		0.030	22.7	−88.8	−66.1	−77.6
5.0		0.060	12.0	−43.7	−31.7	−45.1
6.5	9	0.010	24.2	−84.9	−60.7	−103.1
6.5		0.025	10.1	−35.5	−25.4	−43.7
6.5		0.055	0.2	−16.5	−16.3	−19.5
8.0	7.9	0.010	18.6	−62.4	−43.8	−73.4
8.0		0.025	5.9	−26.3	−20.4	−33.8
8.0		0.055	−1.5	−12.3	−13.8	−16.8

$\beta_{22} \times 10^4$ and contributing terms are given in units of mol ml/g^2. I is the ionic strength (including buffer) in moles/liter. The value z is charge on the protein calculated using intrinsic pK_a values. The sphere diameter for electrostatics calculations, Eq. 5, is 40.0 Å. The contribution from $7\pi a^3/6$ is 7.2×10^{-4} mol ml/g^2.

as such, determines the ionic strength-dependence of β_{22}. Lastly, the contribution from $\beta_{22,s}$ is positive at low ionic strength, but decreases with ionic strength and eventually becomes slightly negative at high ionic strength. These results, combined with those from our earlier study of LYS (22), show the dominance of long-range electrostatic interactions at low ionic strength as well as the importance of short-range electrostatic/dispersion interactions for an accurate description of these protein-protein interactions.

Interestingly, our model of protein-protein interactions for CGA solutions at pH 7.0 shows distinctly different behavior. The results for β_{22} and contributing terms are given in Table 3. Again, comparing the calculated β_{22} with values derived from experiment, we find reasonable agreement and the correct ionic strength-dependence. In this case, however, the positive Donnan contribution approximately offsets the negative contribution from $\beta_{22,l}$ at all ionic strengths, such that β_{22} is determined predominantly by $\beta_{22,s}$. Moreover, $\beta_{22,s}$ is negative at low ionic strength and becomes progressively less negative with increasing ionic strength in contrast to the behavior for LYS and SN. Indeed, we found several CGA-CGA orientations for which local charge effects at contact give rise to short-range attractive electrostatic interactions. It should also be noted that CGA has a net charge of only 4.2 at pH 7.0, approximately half that of LYS or SN at the same pH, which is expected to enhance effects due to its local charge distribution.

Role of specific hydration

Fig. 3 compares β_{22} derived from the SLS experiments to values calculated for SN, LYS, and CGA with and without specific hydration, where the same ensemble of configurations was used in both cases. When specific hydration is considered (Fig. 3 *A*), a linear fit of the data gives a slope ~1.0 with a correlation coefficient of 0.93 and an intercept near zero. Scatter in the data is noticeably larger, however, when specific hydration is not taken into account (Fig. 3 *B*); the correlation coefficient in this case is 0.49. The intercept of ~60 mol ml/g^2 also indicates that removing strongly associated water molecules produces highly complementary configurations resulting in a large negative β_{22}. The β_{22} calculations are also more sensitive to specific hydration at high ionic strength, where short-range dispersion interactions become more important. The net effect is to enhance the influence of strongly associated water molecules in altering the steric complementarity of these interactions, which leads to the nonzero intercept in Fig. 3 *B*. Conversely, the influence of specific hydration on $\beta_{22,s}$ is diminished at low ionic strength, where long-range electrostatic repulsion dominates in the β_{22} calculations. The impact of specific hydration directly on the calculation of osmotic second virial coefficients is shown in Fig. 4, where B_2 for the three proteins calculated with and without specific hydration is compared to experiment. The largest deviations are observed for slightly negative experimental values of B_2, which have significance for protein crystallization (41). In this case, B_2 calculated with specific hydration is in good agreement with

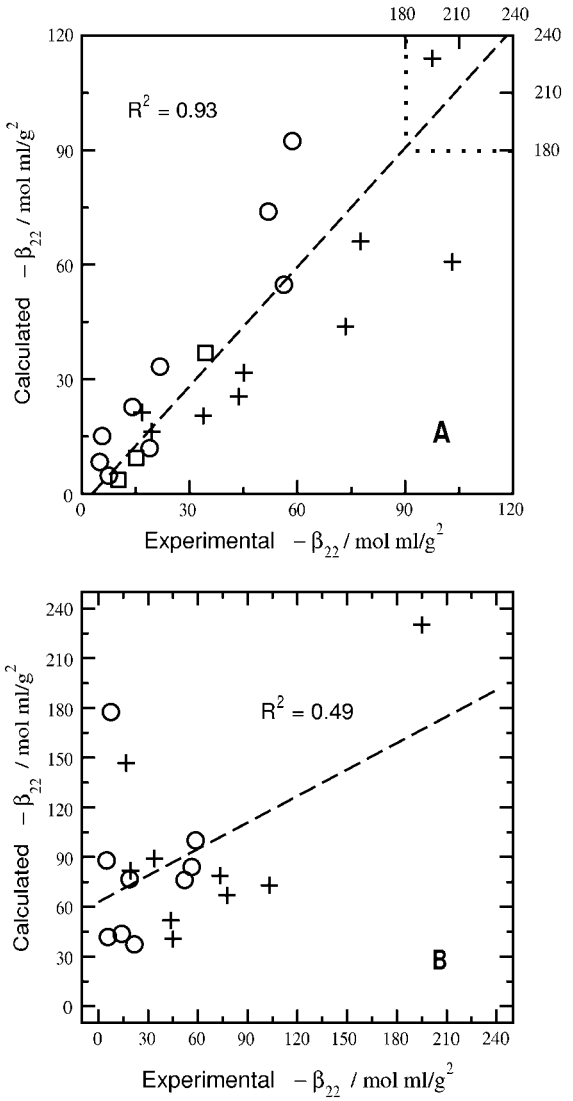

FIGURE 3 Comparison of experimental and calculated values of β_{22} for three globular proteins with (*A*) and without (*B*) bound waters of hydration. Circles, pluses, and squares represent points for LYS, SN, and CGA, respectively. In *B* the calculated values of β_{22} for CGA are on the order of $\sim -10^3$, and therefore, are not shown. The dashed line in *A* is a linear fit of all the points. The dashed line in *B* is a linear fit of the points for SN and LYS.

TABLE 3 Modeling results for chymotrypsinogen at pH 7.0 and charge, $z = 4.2$. $\beta_{22} \times 10^4$ and contributing terms are given in units of mol ml/g^2

I	$\beta_{22,s} + 7\pi a^3/6$	$\beta_{22,l}$	$\beta_{22,calc}$	$2B_2$	$z^2/2\rho_1$	$\beta_{22,exp}$
0.008	−28.5	−8.4	−36.9	−17.6	16.8	−34.4
0.080	−8.3	−1.0	−9.3	−13.4	1.7	−15.1
0.160	−3.1	−0.5	−3.6	−9.4	0.8	−10.2

Experimental data from Velev et al. (6). The sphere diameter for the electrostatics calculations, Eq. 5, is 46.2 Å. The contribution from $7\pi a^3/6$ is 3.4×10^{-4} mol ml/g^2.

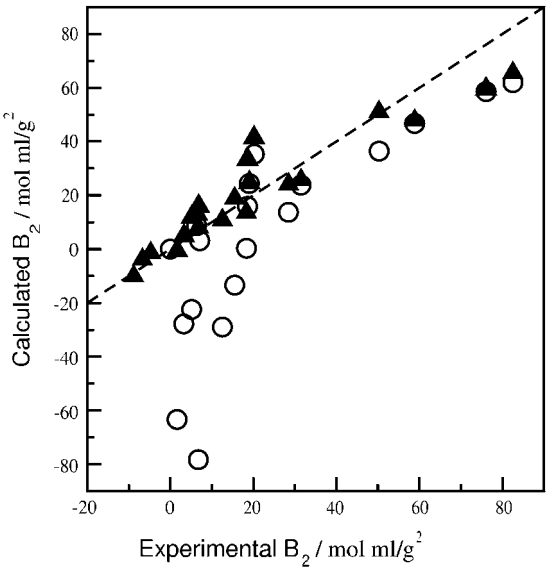

FIGURE 4 Comparison of experimental and calculated values of B_2 for LYS, SN, and CGA with (*solid triangles*) and without (*open circles*) specific hydration. The dashed line is a linear fit of the points with specific hydration and gives a correlation coefficient of 0.94. For experimental values of B_2 less than zero, B_2 calculated without specific hydration is several orders of magnitude more negative than the experimental value, and therefore is not shown.

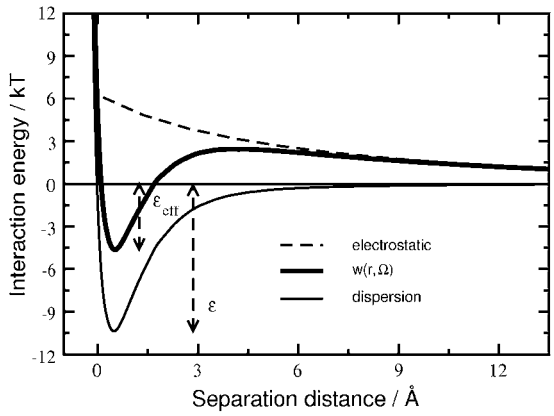

FIGURE 5 Typical profile for PMF, $w(r, \Omega)$, at pH 8.0 and $I = 0.05$ M, between a pair of SN molecules in a given orientation, Ω, as a function of separation distance between them. Profiles for dispersion and electrostatic interaction components of the PMF are also shown. Separation distance is normalized to a value of zero where dispersion interaction energy changes sign. The value $\epsilon(\Omega)$ is the dispersion interaction-energy well-depth. Note that the range of dispersion interactions is much shorter than that of the electrostatic interactions, and the role of repulsive electrostatic interactions in this orientation is effectively to reduce the short-range well-depth from $\epsilon \rightarrow \epsilon_{eff}$.

experiment, whereas the values calculated without specific hydration are more negative by several orders of magnitude.

Although both electrostatic and dispersion interactions contribute to the protein-protein PMF used to calculate $\beta_{22,s}$, the distinguishing feature of this PMF is the short-range nature of the dispersion interactions relative to electrostatic interactions. Fig. 5 shows both interactions calculated as a function of separation distance for a pair of SN molecules in a single, representative orientation at pH 8.0 and $I = 0.05$ M. The dispersion interactions are characterized by an attractive well and a strong repulsion at contact due to the protein-excluded volume. These interactions are clearly short-ranged, and become negligible for separations $>\sim 6$ Å. In contrast, electrostatic interactions—repulsive for this specific orientation—are still significant (~ 1 kT) at a separation of 13.6 Å or one Debye length. They also vary much more slowly with protein-protein separation, and as a result, simply shift up or down the well-depth for dispersion interactions. We take this shifted or effective well-depth, ϵ_{eff}, divided by kT, as a characteristic parameter for describing the short-range interactions embodied in $\beta_{22,s}$.

Fig. 6 compares ϵ_{eff}/kT calculated with and without specific hydration for the same ensemble of SN-SN orientations. Each point in the figure represents one orientation; only those orientations with $\epsilon_{eff}/kT < 0$ are shown. The addition of electrostatic interactions to the dispersion interactions substantially reduces the overall number of favorable protein-protein orientations—i.e., orientations with $\epsilon_{eff} < 0$—such that the number of orientations corresponding to short-range repulsion increases substantially at these solution conditions. Nonetheless, there is a clear bias to more shallow (less attractive) well-depths for those orientations with $\epsilon_{eff} < 0$ when specific hydration is taken into account. The net effect is to preferentially exclude the most attractive SN-SN

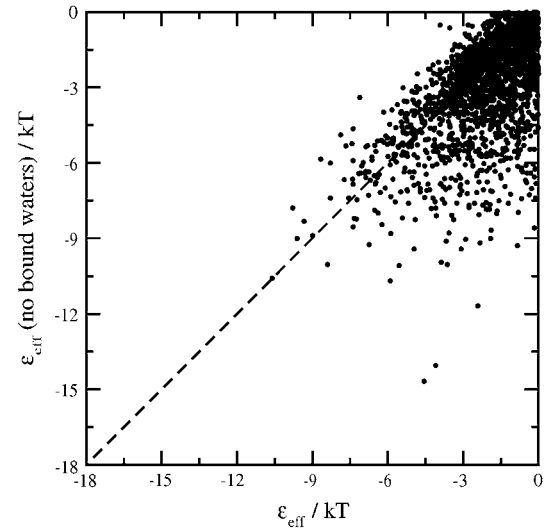

FIGURE 6 Comparison of ϵ_{eff}/kT at pH 8.0 and $I = 0.05$ M for a pair of SN molecules in the same ensemble of protein-protein orientations with and without strongly associated waters of hydration. The calculations were carried out for an ensemble of 10^4 SN-SN orientations; only those orientations with $\epsilon_{eff}/kT < 0$ are shown. The addition of strongly associated waters to SN molecules precludes the most attractive orientations from contributing favorably toward $\beta_{22,s}$, leading to a net increase in $\beta_{22,calc}$ from -146.7×10^{-4} to -13.8×10^{-4} mol ml/g^2 (see Table 2).

orientations from contributing favorably to $\beta_{22,s}$ when strongly associated water molecules are considered, which leads to an order-of-magnitude decrease in the negative value of $\beta_{22,calc}$ from -146.7×10^{-4} to -13.8×10^{-4} mol ml/g^2, in good agreement with the experimental value of -16.8×10^{-4} mol ml/g^2 at pH 8.0 and $I = 0.05$ M (see Table 2).

It should also be noted that the observed discrepancies between the calculated and experimental values of β_{22} (Fig. 3 *B*) cannot be resolved by adjusting the protein-excluded volume to account for steric interactions that arise from a dense hydration layer. Although the additional excluded volume makes a positive contribution to $\beta_{22,calc}$, the increase is much too small to improve agreement significantly with the experimental values of β_{22}. For example, a uniform layer of water 3 Å thick around a protein with a nominal diameter of 40 Å increases the excluded volume contribution to $\beta_{22,calc}$ from 5×10^{-4} to only 7.6×10^{-4} mol ml/g^2. The impact of this contribution is minimal compared to treating specific hydration in terms of its impact on the steric complementarity of short-range protein-protein interactions.

Specific hydration in Fig. 3 is characterized by an ensemble of water molecules that we consider to be strongly associated with these proteins, where the criterion for strong association is $\eta > 2.0$. Here we examine the sensitivity of our results to this value of η. Our approach is to calculate $\beta_{22,s}$, or equivalently $\langle I \rangle_\Omega$ in Eq. 5, as a function of η for dispersion interactions alone. The results for SN are shown in Fig. 7. Decreasing η relaxes the criterion for strong association, and thus corresponds to a larger number of water molecules in the ensemble of strongly associated water molecules. Conversely, increasing η reduces the number of strongly associated water molecules, with $\eta \sim 3.0$ giving $\langle I \rangle_\Omega$ for the protein with no associated waters. Fig. 7 shows that $\langle I \rangle_\Omega$ decreases by an order of magnitude as η is decreased from 3.0 to 2.0, but changes very little for $\eta < 2.0$. Thus, $\eta = 2.0$ defines a lower bound on the number of water molecules that must be considered to obtain the full effect of water association on short-range interactions. Including a larger number of more weakly associated water molecules in our treatment of specific hydration would have only a minimal effect on the $\beta_{22,s}$ calculations.

We also calculated the η distributions for water molecules in the first hydration shell of SN, LYS, and CGA, where η for a particular water molecule is defined by η for the nearest grid point. Fig. 8 shows cumulative fractions of both the first-shell water molecules and the grid points as a function of their assigned η-value for all three proteins. Only those grid points actually occupied by a water molecule during the simulation are included. The most striking observation is that these profiles virtually superimpose despite substantial differences in the three-dimensional structures and spatial charge distributions for these proteins. As expected, the more strongly associated water molecules are also more localized. For example, 60% of the first-shell water molecules have $\eta \leq 1.0$ and occupy nearly 90% of the grid points, whereas the remaining 40% of the water molecules with $\eta > 1.0$ occupy only ~10% of the grid points. Moreover, the strongly associated water molecules ($\eta > 2.0$), which account for ~15% of the total number in the first hydration layer, occupy fewer than 2% of the grid points. The remarkable similarity in hydration behavior for these proteins suggests a universal character to protein hydration when described in terms of the η distribution of first-shell water molecules, consistent with the observation that the fractional compositions of nonpolar, polar, and charged surface regions are similar among diverse proteins (21,42). The results in Fig. 7, on the other hand, indicate that short-range protein interactions are influenced by the specific hydration of only a relatively small fraction of the protein surface; evidently, what differentiates the effect of specific hydration on protein-protein interactions is the variation in the spatial distributions of strongly associated water molecules on the surfaces of these proteins.

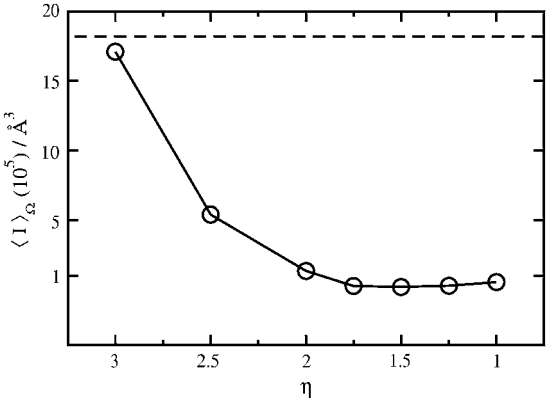

FIGURE 7 $\langle I \rangle_\Omega$ as a function of $\eta = \ln(\rho/\rho_b)$. The dashed horizontal line represents $\langle I \rangle_\Omega$ obtained using the SN molecule with no waters strongly associated to it.

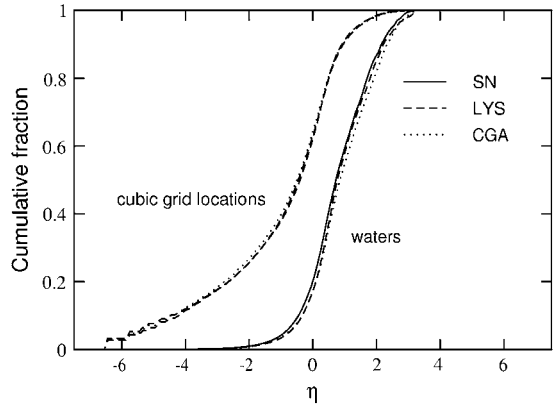

FIGURE 8 The cumulative fraction of cubic grid points (*dashed*) and water molecules (*solid*) present in the first hydration shell of the protein, as a function of $\eta = \ln(\rho/\rho_{bulk})$. Data are shown for three proteins, staphylococcal nuclease (SN), lysozyme (LYS), and chymotrypsinogen (CGA).

CONCLUSIONS

The quasi-chemical description of protein hydration adopted here highlights the important role played by explicit water molecules in mediating weak protein-protein interactions through their local interactions with specific, solvent-accessible regions on the protein surface. Although the thermodynamic affinities that reflect these local interactions have remarkably similar distributions for the three proteins we studied, a key finding of this study is that protein hydration is distinguished by the spatial heterogeneity of strongly associated water molecules on the protein surface and its impact on the surface complementarity of short-range protein-protein interactions. For the three proteins we studied here, strongly associated water molecules identified in molecular dynamics simulations account for $\sim$15% of the total number of water molecules in the primary hydration layer, they are highly localized, and the effect of including them in calculating the interaction part of B_2 is to reduce short-range attractive dispersion interactions by eliminating a number of highly complementary protein-protein contact configurations. Our results lead to the generalization of a finding made specifically in our previous study of B_2 measured by light-scattering from lysozyme solutions; i.e., this treatment of specific hydration is essential for an accurate description of protein-protein interactions embodied in B_2. In contrast, hydration models that consider an effective excluded volume contribution due to a dense hydration layer around the protein cannot resolve discrepancies between our model predictions and experimental data.

Finally, our analysis of protein-protein interactions for staphylococcal nuclease, lysozyme, and chymotrypsinogen as a function of solution pH and ionic strength indicates that differences in the solution thermodynamic behavior and the correlation between B_2 and short-range electrostatic/dispersion interactions that were found to be unique to chymotrypsinogen at certain solution conditions are based on compensating contributions from the Donnan term and long-range electrostatic interactions. This analysis emphasizes the importance of distinguishing the Donnan contribution to B_2, thereby allowing us to interpret the interaction part of B_2 using molecular computations.

We thank Bertrand Garcia-Moreno for access to his laboratory for protein purification and characterization.

Financial support from the National Science Foundation (grants No. BES-0078491 and No. BES-0078844), a Burroughs Wellcome Fund Predoctoral Fellowship for A. Paliwal, and a Howard Hughes Undergraduate Research Fellowship for D. Abras are gratefully acknowledged. We also thank the Ohio Supercomputing Center for a grant of computer time.

REFERENCES

1. Coen, C. J., H. W. Blanch, and J. M. Prausnitz. 1995. Salting out of aqueous proteins: phase equilibria and intermolecular potentials. *AIChE J*. 41:996–1004.
2. Farnum, M., and C. Zukoski. 1999. Effect of glycerol on the interactions and solubility of bovine pancreatic trypsin inhibitor. *Biophys. J*. 76:2716–2726.
3. Haynes, C. A., K. Tamura, H. R. Körfer, H. W. Blanch, and J. M. Prausnitz. 1992. Thermodynamic properties of aqueous α-chymotrypsin solutions from membrane osmometry measurements. *J. Phys. Chem*. 96:905–912.
4. Vilker, V. L., C. K. Colton, and K. A. Smith. 1981. The osmotic pressure of concentrated protein solutions: effect of concentration and pH in saline solutions of bovine serum albumin. *J. Colloid Interface Sci*. 79:548–566.
5. Kirkwood, J. G., and J. B. Shumaker. 1952. Forces between protein molecules in solution arising from fluctuations in proton charge and configuration. *Proc. Natl. Acad. Sci. USA*. 38:863–871.
6. Velev, O. D., E. W. Kaler, and A. M. Lenhoff. 1998. Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen. *Biophys. J*. 75:2682–2697.
7. Eberstein, W., Y. Georgalis, and W. Saenger. 1994. Molecular interactions in crystallizing lysozyme solutions studied by photon correlation spectroscopy. *J. Cryst. Growth*. 143:71–78.
8. Petesev, D. N., and P. G. Vekilov. 2000. Evidence of non-DLVO hydration interactions in solutions of protein Apoferritin. *Phys. Rev. Lett*. 84:1339–1342.
9. Rosenbaum, D., P. C. Zamora, and C. F. Zukoski. 1996. Phase behavior of small attractive colloidal particles. *Phys. Rev. Lett*. 76:150–153.
10. Rosenbaum, D. F., A. Kulkarni, S. Ramakrishnan, and C. F. Zukoski. 1999. Protein interactions and phase behavior: sensitivity to the form of the pair potential. *J. Chem. Phys*. 111:9882–9890.
11. Pauling, L., and M. Delbruck. 1940. The nature of intermolecular forces operative in biological processes. *Science*. 92:77–79.
12. Neal, B. L., D. Asthagiri, and A. M. Lenhoff. 1998. Molecular origins of osmotic second virial coefficients of proteins. *Biophys. J*. 75:2469–2477.
13. Elcock, A. H., and J. A. McCammon. 2001. Calculation of weak protein-protein interactions: the pH dependence of the second virial coefficient. *Biophys. J*. 80:613–625.
14. Svergun, D. I., S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. Protein hydration in solution: experimental observation by x-ray and neutron-scattering. *Proc. Natl. Acad. Sci. USA*. 95:2267–2272.
15. Gerstein, M., and C. Chothia. 1996. Packing at the protein-water interface. *Proc. Natl. Acad. Sci. USA*. 93:10167–10172.
16. Levitt, M., and R. Sharon. 1998. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA*. 85:7557–7561.
17. Merzel, F., and J. C. Smith. 2002. Is the first hydration shell of lysozyme of higher density than water? *Proc. Natl. Acad. Sci. USA*. 99:5378–5383.
18. Liu, W., D. Bratko, J. M. Prausnitz, and H. W. Blanch. 2004. Effect of alcohols on aqueous lysozyme-lysozyme interactions from static light-scattering measurements. *Biophys. Chem*. 107:289–298.
19. Neal, B. L., and A. M. Lenhoff. 1995. Excluded volume contribution to the osmotic second virial coefficient for proteins. *AIChE J*. 41:1010–1014.
20. Henchman, R. H., and J. A. McCammon. 2002. Extracting hydration sites around proteins from explicit water simulations. *J. Comput. Chem*. 23:861–869.
21. Makarov, V., B. M. Pettitt, and M. Feig. 2002. Solvation and hydration of proteins and nucleic acids: a theoretical view of simulations and experiment. *Acc. Chem. Res*. 35:376–384.
22. Asthagiri, D., A. Paliwal, D. Abras, A. M. Lenhoff, and M. E. Paulaitis. 2005. A consistent modeling and experimental approach to light-scattering studies of protein-protein interactions in solution. *Biophys. J*. 88:3300–3309.

23. Stockmayer, W. H. 1950. Light-scattering in multicomponent systems. *J. Chem. Phys.* 18:58–61.

24. Edsall, J. T., H. Edelhoch, R. Lontie, and P. R. Morrison. 1950. Light-scattering in solutions of serum albumin: effects of charge and ionic strength. *J. Am. Chem. Soc.* 72:4641–4656.

25. Cassassa, E. F., and H. Eisenberg. 1964. Thermodynamic analysis of multicomponent solutions. *Adv. Protein Chem.* 19:287–395.

26. Asthagiri, D., B. L. Neal, and A. M. Lenhoff. 1999. Calculation of short-range interactions between proteins. *Biophys. Chem.* 78:219–231.

27. Press, W. H., S. A. Teukolsky, W. T. Vellerling, and B. P. Flannery. 1992. Numerical Recipes in FORTRAN. The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.

28. Paulaitis, M. E., and L. R. Pratt. 2002. Hydration theory for molecular biophysics. *Adv. Protein Chem.* 62:283–310.

29. Wang, J., D. M. Truckses, F. Abildgaard, Z. Dzakula, Z. Zolnai, and J. L. Markley. 1997. Solution structures of staphylococcal nuclease from multidimensional NMR: nucleae-H124L and its ternary complex with Ca^{2+} and Thymidine-3′,5′-Bisphosphate. *J. Biomol. NMR.* 10:143–164.

30. Lamy, A., and J. C. Smith. 1996. Denaturation of truncated staphylococcal nuclease in molecular dynamics simulation at 300 K. *J. Am. Chem. Soc.* 118:7326–7328.

31. Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902–1921.

32. Kalé, L., R. Skeel, M. Bhandarkar, R. Brunner, N. G. N. Kraweta, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.

33. MacKerell, A. D., N. Banavali, and N. Foloppe. 2000. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers.* 56:257–265.

34. Hoover, W. G. 1985. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A.* 31:1695–1697.

35. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald—an $n.log(n)$ method for Ewald sums large systems. *J. Chem. Phys.* 98: 10089–10092.

36. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23: 327–341.

37. Shortle, D., and A. K. Meeker. 1986. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins Struct. Funct. Genet.* 1:81–89.

38. Coumou, D. J. 1969. Apparatus for the measurement of light-scattering in liquids. Measurement of the Rayleigh factor of benzene and some other pure liquids. *J. Colloid Sci.* 15:408–417.

39. Tessier, P. M., S. D. Vandrey, B. W. Berger, R. Pazhianur, S. I. Sandler, and A. M. Lenhoff. 2002. Self-interaction chromatography: a novel screening method for rational protein crystallization. *Acta Crystallogr. D.* 58:1531–1535.

40. Alexandrescu, A. T., E. L. Ulrich, and J. L. Markley. 1989. Hydrogen-1 NMR evidence for three interconverting forms of staphylococcal nuclease: effects of mutations and solution conditions on their distribution. *Biochemistry.* 28:204–211.

41. George, A., and W. W. Wilson. 1994. Predicting protein crystallization from a dilute solution property. *Acta Crystallogr. D.* 50:361–365.

42. Murphy, L. R., N. Matubayasi, V. A. Payne, and R. M. Levy. 1998. Protein hydration and unfolding-insights from experimental partial specific volumes and unfolded protein models. *Fold. Des.* 3:105–118.